J-92-35

NOV 27 1992

SOUTHWEST FISHERIES SCIENCE CENTER

NATIONAL MARINE FISHERIES SERVICE SOUTHWEST FISHERIES SCIENCE CENTER P.O. BOX 271 LA JOLLA, CA 92038

OCTOBER 1992

USING VERTICAL AERIAL PHOTOGRAPHS TO ESTIMATE DOLPHIN SCHOOL SIZES: PRECISION AND CONSISTENCY

LIBRARY

JAN 11 1993

National Marine Fisheries Service
2570 Dole Street
Honolulu, Hawaii

By

James W. Gilpatrick, Jr.

ADMINISTRATIVE REPORT LJ-92-35

This Administrative Report is issued as an informal document to ensure prompt dissemination of preliminary results, interim reports and special studies. We recommend that it not be abstracted or cited.

USING VERTICAL AERIAL PHOTOGRAPHS TO ESTIMATE DOLPHIN SCHOOL SIZES: PRECISION AND CONSISTENCY

James W. Gilpatrick, Jr.
National Oceanic and Atmospheric Administration
National Marine Fisheries Service
Southwest Fisheries Science Center
8604 La Jolla Shores Drive
La Jolla, CA 92038

October 1992

Administrative Report LJ-92-35

ABSTRACT

To improve abundance estimates of dolphin populations in the eastern tropical Pacific, dolphin counts derived from aerial photographs are used to "calibrate" visual estimates of school size made by observers on board a research vessel. Because there is a degree of subjectivity in counting dolphin images on photographs, methods are described that minimize variability in photo-counts. Photographed dolphins are counted by three independent readers during several consecutive months annually. To evaluate how reliable the methods were and if criteria used for counting were applied consistently by the readers during these months, temporal trends in estimates were tested by linear regression analyses and by way of an experiment using repeated measures analysis of variance.

The coefficient of variation (CV) for replicate estimates of 48 schools averaged 5.39% and ranged between 1.24% and 14.58%. Most (92%) were estimated with precision of 9.0% CV or better. The average percent error for readers ranged between 3.51% and 4.67%, indicating that readers were quite precise, and regression analyses indicated no temporal trends or bias in the distribution of counts relative to the means. The repeated measures experiment showed a significant "reader with time" effect attributable to the higher variability of counts made at the beginning of the experiment. However, the contribution of this variance component to total variability with time proved insignificant. These results suggested that methods were reliable, and readers were precise and consistent after an initial learning period.

Digitized by the Internet Archive
in 2025

https://archive.org/details/administrative-report-lj_1992-10_35

INTRODUCTION

Accurate population estimates of dolphins are dependent on accurate estimates of the number of individuals in a school. School size estimates are often made by observers on board vessels as they conduct line transect surveys for dolphins (Wade and Gerrodette, 1992). Accurate estimation of the number of animals seen in a group is difficult, and group size is often underestimated by observers (Erwin, 1982; Scott et al., 1985). Estimating dolphin school sizes from a ship is especially difficult because, due to the diving behavior of the dolphins, the entire school is rarely visible at the sea surface at one time. As a result, school size estimates can be highly variable, and they may be biased (Anganuzzi and Buckland, 1989). If this is true, then population abundance estimates derived from these data may be biased and resulting management decisions aimed at conserving these populations may be inappropriate.

Taking aerial photographs of a dolphin school simultaneously as the shipboard observers are estimating its size is a means of acquiring precise estimates and determining whether the observer's estimates are biased. Photographic estimates derived from large format aerial cameras used at the Southwest Fisheries Science Center (SWFSC) were validated during a 1979 investigation where five separate dolphin schools (size range: 161 - 396 dolphins) were photographed, then captured in a tuna purse-seine net. Results showed that counts from the photographs were no different than counts tallied by hand as dolphins were being released from the net (Scott et al., 1985).

The SWFSC has been monitoring eastern tropical Pacific (ETP) populations of the pantropical spotted (*Stenella attenuata*), spinner (*S. longirostris*), striped (*S. coeruleoalba*) and common (*Delphinus delphis*) dolphins because many have been killed in the nets of the multinational purse-seine fishery for tunas (Perrin, 1975). Starting in 1987, annual population surveys were complemented with photographic sampling from a ship-based helicopter so that shipboard observer estimates could be compared and "calibrated" with photo estimates of the same schools.[1] However, this process assumes that dolphin counts made from the photographs are themselves accurate and precise.

This report details photographic and counting methods used at the SWFSC to derive dolphin school size estimates from large-format vertical aerial photographs. Many of the techniques utilized originated with Scott et al. (1985); the techniques have been modified here for application to large-scale, annual survey efforts designed to estimate dolphin populations in the ETP (i.e., annual ocean surveys of 120 days duration each covering a 19 million km² study area). The precision of replicate count estimates for 48 dolphin[2] schools photographed during a 1989 dolphin survey are also reported. In a typical year, more than 200 photographs of dolphin

---

[1]Gerrodette, T. and C. Perrin. 1991. Calibration of shipboard estimates of dolphin school size from aerial photographs. Admin. Rpt. LJ-91-36, 18 p.; available from Southwest Fisheries Science Center La Jolla Laboratory, National Marine Fisheries Service, NOAA, P.O. Box 271, La Jolla, CA 92038.

[2]For the purposes of this paper, "dolphin" refers to dolphins as well as small toothed whales that are included in the photograph sample.

schools are counted by three independent readers over several consecutive months. One concern was that during these "counting periods", readers might be inconsistent in the application of cognitive and/or physical methods used for counting and this might effect the accuracy and precision of the photo estimates. To evaluate if counting methods were reliable and if readers were consistent in the application of criteria used for counting, temporal trends in estimates were examined.

MATERIALS AND METHODS

At sea

Environmental conditions were optimal for photography when sun angle was high enough to allow sufficient light penetration of the sea surface and when sea state was calm enough, usually ≤ Beaufort 3, to minimize sun-glare reflection which can obscure dolphins in the photographs. Aerial photographs were taken from a Hughes 500-D helicopter deployed from the National Oceanic and Atmospheric Administration (NOAA) ship *David Starr Jordan*. All cameras (Chicago Aerial Industries KA-62 model) were vertically mounted, had forward-motion compensation, and a 76.2-mm lens. Large-format (126-mm) Kodak Aerochrome MS 2448 color film was used. In order to minimize the dolphin schools behavioral response (scattering and deep diving) to the helicopter, photos were taken at a typical altitude of 244 m. The scale of the photographs at this altitude was 1:3200, the sea surface area in a photo frame was 366 m^2, and a 2 m dolphin measured .625 mm on the film.

The camera cycle rate was programmed to expose for approximately 80% film image overlap from frame-to-frame, i.e., 80% of the area photographed in one frame was photographed again in the next successive frame; this enabled readers to track individual dolphins for up to four consecutive frames (Fig. 1). Successive exposed frames over a school were recorded as a complete "photo-pass." To enhance the probability of capturing the entire school in a photo-pass, multiple photo-passes were made over a school. If some (usually only a few) dolphins were outside of the area photographed, they were noted by the aerial photographers and the count was later added to the estimate of school size.

In the photography laboratory

Color photograph transparencies were processed and reviewed for image quality prior to counting. Passes where dolphins were noted as missed or swimming too deep to be clearly seen were eliminated from the sample. Images were counted using Richards Corp., GFL-940 MCE type light tables equipped with dissection microscopes having .7x to 7x variable objective and 10x wide-field oculars.

Dolphins were counted by hand-tally while being plotted with permanent marker on a clear acetate overlay. The marked overlay was then moved to the image overlap area of adjacent frames and aligned over dolphins that were previously plotted. This allowed readers to identify dolphins not previously counted and to re-examine

"questionable images" between adjoining frames (consistent with method used by Scott et al. 1985). The photo-pass was the unit on which school size estimates were based; each pass was counted independently by three readers, and estimates were not discussed among readers.

Quality ratings

After being counted, each pass was assigned a subjective "quality rating" by the readers. Ratings reflected how confident the reader was in the accuracy of the count; this was based on the level of difficulty with which the school was counted. The level was dictated by image resolution and the behavior and configuration of the school. For instance, a photo-pass where dolphin images were blurred because dolphins were densely grouped and swimming deep below the surface would have more questionable images, would be more difficult to count, and would convey less confidence than a pass where dolphins were spaced out and clearly photographed at the surface. A rating of "1" indicated that a photo-pass had "excellent" quality and dolphins were easy to count. A rating of "2" indicated that the count was "good" despite the presence of some questionable images. A photo-pass was rated "3" or "fair" when more questionable images were encountered than a "2" rated pass but, readers still believed the count was a close approximation of true school size. A pass rated "4" was deemed unusable for school size estimation because the photos could not be interpreted with an acceptable degree of confidence.

Sources of variance

Variability between repetitive counts was attributed to "counting error," where dolphins were missed or double-counted by a reader. Variability also occurred when partially obscured images were interpreted differently between readers. This was a subjective source of variance because one reader might decide that a deep, barely visible image was a dolphin while another reader may have viewed it as merely background water turbulence in the photo.

To monitor variability, count data were reviewed to identify passes with extreme outlying counts or passes with a coefficient of variation (CV) greater than 10%. These passes were scrutinized for counting error, and often the pass was recounted. To avoid bias in repeated counts, readers were not told whether their original count fell high or low with respect to the other counts. If, after a repeat count, the CV remained greater than 10%, the field and lab notes and data were reviewed by the group, and a consensus decision was made concerning the accuracy of the photo-pass. Photo-passes with a CV greater than 15% were typically eliminated from the sample.

The "true" school size

If multiple photo-passes were made for a single school, typically, the pass with the highest mean of the three independent counts and the best (i.e., lowest numeric) quality ratings was selected as the best or "true" estimate for the school. In all other cases (usually very few), where this relationship differed

(i.e., the highest or lowest mean estimate was accompanied by, respectively, the highest and lowest quality ratings), the notes and data were reviewed, and the most accurate pass was designated by group consensus.

Analytical methods

Descriptive indices of precision

The CV statistic was used to screen data (noted above) and also as an overall index of between- and within-reader precision. Additionally, because quality ratings are considered when selecting the most accurate photo-passes, the CV was used in evaluating the relationship between estimate precision and quality ratings. The average percent error (APE; described by Beamish and Fournier, 1981) index was used to evaluate "reader" precision by describing how far, on average, the readers' counts were from their respective means. A low reader APE value would indicate that a reader was precise because counts were close to their respective means. It should be noted that the CV and APE are different: the CV describes precision using the standard deviation in proportion to (divided by) the mean while the APE uses the absolute value of the difference between a particular estimate (a readers' count) of school size and the mean estimate. Given the same data set for analyses, the CV is greater than the APE by a consistent factor (Chang, 1982). When viewed separately, however, the CV and APE are consistent indices of precision relative to the mean.

Additionally, the APE equation was modified here to evaluate, for each independent count, the absolute value (referred to here as percent error = PE) and actual value (referred to as percent deviation = PD) of its distance from the mean.

For the APE, PE, and PD indices, let x_{ij} be the ith reader's determination of the size of school j. The average determination over three readers of the size of the jth school is given by

$$\bar{x}_{.j}=\frac{1}{3}\sum_{i=1}^{3} x_{ij} \qquad (1)$$

The PE of x_{ij} is expressed as

$$PE_{ij}=\frac{|x_{ij}-\bar{x}_{.j}|}{\bar{x}_{.j}}*100 \qquad (2)$$

The APE then, is the average difference between reader i's counts and the respective mean estimates for the sample of N schools expressed as

$$APE_i=\frac{1}{N}\sum_{j=1}^{N}\frac{|x_{ij}-\bar{x}_{.j}|}{\bar{x}_{.j}}*100 \qquad (3)$$

To examine the distribution of counts relative to the means or true school sizes, PD values were computed and plotted. PD is essentially the deviate statistic, which measures how distant, in a negative or positive direction, a count is from the mean (Sokal

and Rohlf, 1981), expressed as a percentage of the mean. The PD of x_{ij} is expressed as

$$PD_{ij}=\frac{x_{ij}-\bar{x}_{.j}}{\bar{x}_{.j}}*100 \qquad (4)$$

The frequency distribution of PD values were plotted to evaluate if individual reader counts were normally distributed or if readers tended to count high or low relative to the means. To evaluate if there was a temporal trend in individual reader count deviation during the 80-day counting period, PD values were regressed against the variable "time." In this case, "time" represented the chronological sequence, unique for each reader, in which the 48 schools were counted. Logistically it was impossible for readers to follow the same sequence in counting photographs; this resulted in schools being counted more or less at random between readers during the period.

Repeated measures experiment

To test the validity of counting methods and whether readers were consistent in criteria and methods applied during the 80-day counting period, a repeated measures experiment was conducted. A standard sample of six schools (two pantropical spotted, two spinner, and two striped dolphin schools), which varied in size and image quality, were counted repeatedly during the counting period so that changes with time in estimates attributable to the

variables school size, image quality, and reader could be detected.

To provide a reasonably even distribution of school sizes among the standard sample, schools were selected from a large sample (photographed during 1987 and 1988) categorized into small schools (< 120 dolphins), medium schools (from 120 to 350 dolphins), and large schools (> 350 dolphins) (Fig. 2). To detect variability due to image quality, two schools, one with an average quality rating of "1" or "excellent" and another with an average of "2" or "good," were selected within each of the small- and medium-size categories. For the large-size category, the two available schools each had average ratings of 1.33. Because this was close to "1," the passes were entered into the model under the "excellent" rating category. Variability due to individual readers was tested by categorizing counts according to the respective readers at levels one, two, and three. It was assumed that there was no effect attributable to species differences among the counts because differences in color morphology and modal differences in length distributions between the dolphin species are not visually noticeable at the scale used for counting.

Counts of the standard sample were repeated four times: once at the start of the counting period then again every 25 days during the period. To minimize the possibility that readers would recognize the same photos during the experiment, no catalog numbers were evident for the photos and they were counted at random and interspersed with the regular counting period photos.

Repeated counts for a pass were considered the "within-

treatment" observations or data for that school. Within-treatment data were reviewed for outliers by visual screening and using the Shapiro-Wilk's test for normality with significance at $\alpha = .05$. Six outliers were excluded from the analysis because they were departures from expected observations. After log-transformation, data met F-test requirements of homoscedasticity according to Levene's test at $\alpha = 0.05$ (computer program BMDP7D used, Dixon et al. 1988). Transformed data were found to be normally distributed (BMDP1R, Dixon et al., 1988). The profile of within-treatment data (categorized into levels of the various factors) over time and differences between factors were tested with a repeated measures analysis of variance (ANOVA) model modified from Winer (1971, p. 337) of the following form:

$$X_{ijkl(m)} = \mu + \alpha_i + \beta_j + \delta_k + \pi_{m(ijk)} + \tau_l + \alpha\tau_{il} + \beta\tau_{jl} + \delta\tau_{kl} + \tau\pi_{lm(ijk)}$$

In the model, all factors are assumed fixed; the various parts are

μ = the overall mean,

α_i = the between-factor school size (with three levels),

β_j = the between-factor quality rating (with two levels),

δ_k = the between-factor reader (with three levels),

$\pi_{m(ijk)}$ = error term for between-factors,

τ_l = the within-factor time (with four levels),

$\alpha\tau_{il}$ = change in the factor school size with time,

$\beta\tau_{jl}$ = change in the factor quality rating with time,

$\delta\tau_{kl}$ = change in the factor reader with time, and

$\tau\pi_{lm(ijk)}$ = error term for the interaction of factors with time.

After removal of outliers, data were equally weighted in hypotheses testing by use of the ANOVA model with type III sum of squares (Milliken and Johnson, 1984). To adjust for missing observations, the F-test degrees of freedom for the residual mean square were reduced by one for each missing value. A 0.05 level of significance was used for the ANOVA. The ANOVA was computed using software program Abacus Concepts, SuperANOVA™ (1990).

In using the model, it was necessary to assume that the data were symmetrical (i.e., the observations between any two experiment time points were not overly correlated; see Dunn and Clark, 1987 and Gurevitch and Chester, 1986). Although the test by Anderson (1958) showed the correlation in the data was not significant ($p = .3246$ at $\alpha = .05$), conservative adjustments of the *df* for the within-factor F-tests were made using the Greenhouse-Geisser (1959) epsilon factor of 0.692. This adjusted the significance of the F-tests to nominal levels based on the amount of correlation in the data (Barcikowski and Robey, 1984).

RESULTS

Counting period

Replicate count estimates for 48 dolphin schools photographed in 1989 were made independently by three readers during the 80-day counting period. School size estimates ranged between 4 and 633 with an average of 146 (values rounded off to nearest whole numbers; Table 1). The precision of estimates ranged from 1.24% to 14.58% CV with an average of 5.39% (Table 1). Most of the schools

(92%) were estimated with precision better than 9.0% CV (Fig. 3). There was little difference in precision between small, medium, and large school estimates (Table 2). Count precision tended to decrease with decreased quality of the images (Fig. 4).

The PE of individual reader counts ranged from 0 to 16.79%; reader two counted most precisely during the period with an APE equal to 3.51%. Readers one and three followed with APE values of 3.75% and 4.67% respectively (Table 1). The frequency distribution of PD values, although not identical between readers (Fig. 5), appeared to be normally distributed about the means. PD values regressed on "time" appeared randomly distributed about the means and showed no significant temporal trends for reader counts during the counting period (Fig. 6).

Repeated measures experiment

Repeated counts for the standard sample are presented in Table 3. These schools were estimated with precision comparable with that of "most" of the counting period schools (Table 2 and 3). Differences in the range of precision between the two data sets were most likely attributable to the small sample (n = 6) used in the experiment. Reader two was most consistent in repeated counts of the same schools (average CV = 2.58%) followed by readers three and one (average CV = 3.38% and 4.67% respectively; Table 4).

The repeated measures ANOVA of transformed data (Table 5) showed significant differences in means for the between-factors school size and quality rating. This was not surprising because

different sized schools were selected for the analysis. Often, with repeated measures ANOVA, between-factor differences are confounded (Winer, 1971). There were no differences between the overall means of the standard sample counts for each reader.

The temporal trends of the within-factor data were of primary interest in the repeated measures ANOVA. The F value for the within-factor "time" indicated no significant change for the treatment means across the four time points. There was also no change in estimates attributed to school size and quality ratings. A significant interaction for "reader with time" (F= 3.503, p = .0258, df = 6, 31) indicated that the mean count profiles for each reader differed during the experiment (Fig. 7). Figure 7 suggested that the mean for reader three's counts made at time four was high and the mean of reader one's counts made at time one was low relative to their respective time profiles. These differences were tested *a posteriori* by contrasts among means using linear coefficients to weigh the log-transformed data (Sokal and Rohlf, 1981; Abacus Concepts, SuperANOVA™, 1989). For counts done by reader three, the mean of counts made at time four did not differ significantly from the mean of counts made at times one, two, and three (F = 3.535, p = .0695, df = 1). For reader one, the mean of counts made at time one differed significantly from the mean of counts made at times two, three, and four (F = 14.977, p = .0005, df = 1). This indicated that reader one's counts made at time one were important in causing the significant "reader with time" effect.

DISCUSSION

Precision and Consistency

Scott et al. (1985) reported that dolphin school size estimates derived from aerial photographs were accurate and more precise than visual estimates. They found the standard deviation of estimates (log transformed) averaged 6% of school size for photographic estimates and 10% - 30% of school size for visual estimates. The CV for estimates (untransformed) of the 11 schools used in their precision analysis averaged 8.39%[3] indicating slightly less precision when compared to estimates presented here (avg. 5.39% CV). The difference is explained, in part, by their statistical model which accounted for variance due not only to independent repetitive counts, but also due to camera types (126 and 229 mm formats) and multiple photo-passes (2 to 7) for a given school. In the present analysis, variability was minimized by use of one type of camera (126mm format) and by including only estimates of the single best photo-pass for a school.

Results presented here indicate that readers were quite precise in counting and that no bias or temporal trends in individual reader counts of the photographs were detected. Reader two was most experienced in counting and was most precise and consistent according to the indices. Figure 4 showed there was a tendency for reader precision to decrease with decreasing photo

---

[3]Clark, W.G. 1984. Analysis of variance of photographic and visual estimates of dolphin school size. Admin. Rpt. LJ-84-11C, 18 p.; available from the Southwest Fisheries Science Center La Jolla Laboratory, National Marine Fisheries Service, NOAA, P.O. Box 271, La Jolla, CA 92038.

image quality. This relationship, however, should be viewed in a general way when ratings are considered in selecting the most accurate or "best" photo-passes because there are considerable variation in these data. This probably resulted from the degree of subjectivity associated with the quality rating process.

Repeated measures experiment

The experiment demonstrated that comparable and precise dolphin school size estimates can be derived repeatedly from aerial photographs taken and interpreted with the SWFSC system. There was no significant trend in the overall means of counts over time despite a significant difference between readers with time. The significant effect was attributable to the higher variability of the counts made at the beginning of the experiment. Because the initial experiment counts were made just prior to the beginning of the counting period, this suggested that variance was atypically high early in the counting period as well. The regression analyses of PD values with time, however, did not provide strong evidence for this because counts appeared precise and normally distributed about the means. Probable cause for the significant effect was that readers were "out of practice" at the start of the experiment. Before starting the 1991 counting period, a practice or "learning" counting session (now, a standard annual procedure) was effective in bringing the between-reader variance down from an average of about 8% to 4% CV within one day. Results suggested that after a learning period, readers were consistent in criteria applied while

counting.

Counting technique and outlier values

When dolphins were densely aggregated in a portion of the school, the most efficient way to estimate the school was to initiate plotting and counting at the high density area of the pass and then continue to plot and count those dolphins photographed on the adjoining frames. This reduced the likelihood of missing or double-counting dolphins because most of the school was plotted and counted initially making it easier to track dolphins found on adjoining frames. Outlier values for standard sample school number III (Table 3) resulted because dolphins were missed when readers started counting in the dolphin low density area of the pass (in this case the beginning of the pass) then continued plotting and counting into the dolphin high density area of the pass.

CONCLUSION

The field and laboratory methods described proved reliable in providing precise school size estimates for large scale efforts to estimate dolphin populations in the ETP. Results from analyses of both the counting period and experiment standard sample counts provided complementary information on the precision, consistency, and some of the sources of variance in photo estimates. The indices of precision applied here will be used for inter- and intra-annual quality comparisons and monitoring of dolphin school photographic

estimates. The analytical methods used in this paper to describe precision and consistency are applicable in other fisheries and terrestrial wildlife research efforts where independent repetitive counts are conducted for parameter estimation.

Acknowledgements

Thanks to Drs. N. Chyan-huei Lo, SWFSC and Boyd Collier, San Diego State University for suggestions on analyses. Thanks to fellow members of the Photogrammetry and Remote Sensing Group at the SWFSC: W. L. Perryman, R. L. Westlake and M. S. Lynn. R. M. Allen, H. M. Orr and K. Raymond drafted figures and tables. Dr. T. Gerrodette reviewed an earlier version of the manuscript. W. L. Perryman, C. A. M. Salvado', Drs. D. P. DeMaster, A. Dizon, M. D. Scott, J. Barlow and S. B. Reilly also reviewed the manuscript and offered useful suggestions. C. Blair assisted with editing.

CITATIONS

Abacus Concepts, SuperANOVA™. 1990. SuperAnova: Accessible General Linear Modeling. Abacus Concepts, Inc., Berkeley, California, 322 p.

Anderson, T.W. 1958. An Introduction to Multivariate Statistical Analysis. John Wiley & Sons, Inc., New York. 374 p.

Anganuzzi, A.A. and S.T. Buckland. 1989. Reducing bias in trends in dolphin relative abundance, estimated from tuna vessel data. Rep. Int. Whal. Commn. 39:323-334.

Beamish, R.J. and D.A. Fournier 1981. A method of comparing the precision of a set of age determinations. Can. J. Aquat. Sci. 38:982-983.

Barcikowski, R.S. and R. R. Robey. 1984. Decisions in single group repeated measures analysis: Statistical test and three computer packages. Amer. Statistician 38(2):148-150.

Chang, W.Y.B. 1982. A statistical method for evaluating the reproducibility of age determination. Can. J. Fish. Aquat. Sci. 39:1208-1210.

Dixon, W. J., M.J. Brown, L. Engleman, M.A. Hill and R.I. Jennrich. 1988. BMDP Statistical Software Manual. Univ. Calif. Press, Berkeley, 1234 p.

Dunn, J. D and V. A. Clark 1987. Applied Statistics: Analysis of Variance and Regression. Second edition. John Wiley & Sons, Inc., New York, 445 p.

Erwin, R. M. 1982. Observer variability in estimating numbers: an experiment. J. Field Ornithol. 53:159-167.

Greenhouse, S. W. and S. Geisser. 1959. On methods in the analysis of profile data. Psychometrika 24:95-112.

Gurevitch, J. and S. T. Chester, Jr. 1986. Analysis of repeated measures experiments. Ecology 67:251-255.

Milliken, G. A. and D. E. Johnson. 1984. Analysis of Messy Data, Vol. 1: Designed Experiments. Lifetime Learning Publ., Belmont, California, 473 p.

Perrin, W.F. 1975. Distribution and differentiation of populations of dolphins of the genus *Stenella* in the Eastern Tropical Pacific. J. Fish. Res. Board Can. 32:1059-1067.

Scott, M. D., W.L. Perryman and W. G. Clark. 1985. The use of aerial photographs for estimating school sizes of cetaceans. Inter-Amer. Trop. Tuna Comm. Bull. 18(5):383-404 [In Engl. and Span.].

Sokal, R. R. and F. J. Rohlf. 1981. Biometry: The Principles and Practice of Statistics in Biological Research. Second edition. W.H. Freeman and Co., New York. 859 p.

Wade. P. R. and T. Gerrodette. 1992. Estimates of dolphin abundance in the eastern tropical Pacific: Preliminary analysis of five years of data. Rep. Int. Whal. Commn. 42:533-546.

Winer, B. J. 1971. Statistical principles in experimental design. Second edition. McGraw-Hill, New York.

Table 1

ts, mean school size, indices of precision and average quality ratings schools photographed in the 1989 ETP dolphin survey.

ol	Species codes[a]	Reader counts[b]			Mean school size	CV (%)	Avg. quality[c] rating	Reader PE[d]		
		R1	R2	R3				R1	R2	R3
	S.COE	37	35	39	37.00	5.41	1.33	0.00	5.41	5.41
	S.ATT	250	227	221	232.67	6.58	2.66	7.45	2.44	5.01
	S.LON	153	139	147	146.33	4.80	1.00	4.56	5.01	0.46
	S.COE	57	54	56	55.67	2.74	2.33	2.40	2.99	0.60
	S.ATT/S.LON	124	127	130	127.00	2.36	2.33	2.36	0.00	2.36
	S.ATT/S.LON	192	196	216	201.33	6.39	1.66	4.64	2.65	7.28
	S.ATT/S.LON	274	236	242	250.67	8.15	1.00	9.31	5.85	3.46
	S.COE	48	47	49	48.00	2.08	1.00	0.00	2.08	2.08
	F.ATT	17	18	18	17.67	3.27	1.00	3.77	1.89	1.89
	S.ATT/S.LON	81	78	77	78.67	2.65	1.00	2.97	0.85	2.12
	S.ATT/S.LON	412	416	398	408.67	2.31	1.33	0.82	1.79	2.61
	S.ATT/S.LON	101	106	118	108.33	8.06	2.33	6.77	2.15	8.92
	S.LON	20	20	21	20.33	2.84	2.66	1.64	1.64	3.28
	S.ATT/S.LON	107	113	113	111.00	3.12	1.33	3.60	1.80	1.80
	S.ATT	118	124	109	117.00	6.45	1.33	0.85	5.98	6.84
	S.ATT/S.LON	618	607	675	633.33	5.76	2.00	2.42	4.16	6.58
	P.ELE	400	391	399	396.67	1.24	1.00	0.84	1.43	0.59
	S.COE	58	53	60	57.00	6.33	1.33	1.75	7.02	5.26
	S.COE	52	55	61	56.00	8.18	1.66	7.14	1.79	8.93
	D.DEL	317	323	326	322.00	1.42	1.00	1.55	0.31	1.24
	D.DEL	326	312	367	335.00	8.53	1.33	2.69	6.87	9.55
	S.COE	24	23	25	24.00	4.17	1.33	0.00	4.17	4.17
	G.MAC	19	18	19	18.67	3.09	1.00	1.79	3.57	1.79
	S.COE	75	77	68	73.33	6.44	2.00	2.27	5.00	7.27
	S.COE	34	30	32	32.00	6.25	1.33	6.25	6.25	0.00
	S.COE	166	171	181	172.67	4.42	1.66	3.86	0.97	4.83
	D.DEL	66	64	60	63.33	4.82	1.66	4.21	1.05	5.26
	S.ATT/S.LON	216	216	233	221.67	4.43	1.33	2.56	2.56	5.11
	S.ATT	25	24	23	24.00	4.17	1.00	4.17	0.00	4.17
	G.GRI	56	64	55	58.33	8.46	1.33	4.00	9.71	5.71
	S.BRE	4	4	5	4.33	13.32	2.00	7.69	7.69	15.38
	S.LON	576	678	548	600.67	11.39	2.33	4.11	12.87	8.77
	S.ATT	38	50	49	45.67	14.58	2.66	16.79	9.49	7.30
	S.COE	40	39	40	39.67	1.46	1.00	0.84	1.68	0.84
	S.ATT	88	90	80	86.00	6.15	1.66	2.33	4.65	6.98
	S.LON	315	324	350	329.67	5.51	1.33	4.45	1.72	6.17
	S.COE	26	25	24	25.00	4.00	2.00	4.00	0.00	4.00
	S.COE	150	147	156	151.00	3.03	2.00	0.66	2.65	3.31
	S.BRE	36	34	35	35.00	2.86	1.00	2.86	2.86	0.00
	S.ATT/T.TRU	284	293	337	304.67	9.31	2.33	6.78	3.83	10.61
	G.GRI	20	19	18	19.00	5.26	1.00	5.26	0.00	5.26
	S.COE	35	34	35	34.67	1.67	1.00	0.96	1.92	0.96
	S.COE	73	74	88	78.33	10.71	2.00	6.81	5.53	12.34
	S.COE	154	154	144	150.67	3.83	1.33	2.21	2.21	4.42
	D.DEL	450	526	474	483.33	8.04	1.33	6.90	8.83	1.93
	S.COE	80	87	92	86.33	6.98	1.66	7.34	0.77	6.56
	S.COE	23	24	23	23.33	2.47	1.33	1.43	2.86	1.43
	S.COE	39	39	37	38.33	3.01	2.00	1.74	1.74	3.74
verage:		145.5	145.9	147.4	145.50	5.39	APE[e]=	3.75	3.51	4.67

Table 1 (continued)

---

cies codes: S.COE = *Stenella coeruleoalba*; S.ATT = *S. attenuata*; S.LON = *ongirostris*; F.ATT = *Feresa attenuata*; P.ELE = *Peponocephala electra*; C = *Globicephala macrorhynchus*; D.DEL = *Delphinus delphis*; G.GRI = *pus griseus*; S.BRE = *Steno bredanensis*; T.TRU = *Tursiops truncatus*. iple species codes indicate mixed-species schools.

der codes: R1 = reader one; R2 = reader two; R3 = reader three.

ained by averaging the three independent reader assigned quality ratings.

= percent error.

= average percent error (Beamish and Fournier, 1981).

Table 2

Comparison of precision of estimates for all small (< 125 dolphins), medium (between 125 - 350 dolphins), and large (> 350 dolphins) schools counted during the counting period and the experiment.

School Size		Counting Period CV (%)	Repeated Measures Experiment CV (%)
Small	mean	5.3	4.3
	range	(1.45-14.58)	(1.54-6.25)
Medium	mean	5.2	4.7
	range	(1.42-9.31)	(.75-4.94)
Large	mean	5.7	4.4
	range	(1.24-11.39)	(.98-7.66)
overall mean :		5.4	4.5
overall range:		(1.24-14.58)	(.75-7.66)

Table 3.

Repeated measures experiment counts. Counts were repeated at time points which were 25 days apart during the 80 day counting period. Observations in parentheses are outlier values.

SCHOOLS	SCHOOL SIZE	QUALITY RATING	READER	TIME			
				TIME 1	TIME 2	TIME 3	TIME 4
I	SMALL	EXCELLENT	1	64	64	66	67
			2	69	66	68	70
			3	61	65	64	69
II	SMALL	GOOD	1	34	38	38	35
			2	36	35	36	36
			3	38	39	39	38
III	MEDIUM	EXCELLENT	1	(191)	242	230	236
			2	232	252	236	243
			3	(189)	(193)	(205)	(206)
IV	MEDIUM	GOOD	1	(169)	214	194	203
			2	207	211	209	206
			3	213	195	194	205
V	LARGE	EXCELLENT	1	401	468	468	450
			2	457	450	429	432
			3	440	451	429	467
VI	LARGE	EXCELLENT	1	464	516	525	490
			2	541	527	534	496
			3	507	519	534	527

Table 4.

Individual reader precision in counting the same schools repeatedly.

Reader	Coefficient of Variation	
	Average (%)	Range (%)
1	4.67	(2.54 - 7.08)
2	2.58	(1.39 - 3.78)
3	3.38	(1.50 - 5.10)

Table 5.

Results from repeated measures analysis of variance (ANOVA).

	df	Sum of Squares	Mean Square	F-Value	P-Value	G.-G.[1] P-Value
School Size	2	8.880	4.4400000	431.485	0.0001	
Quality Rating	1	0.290	0.2900000	28.145	0.0003	
Reader	2	0.005	0.0020000	0.241	0.7898	
Error (Between factors)	11	0.113	0.0100000			
Time	3	0.002	0.0010000	2.504	0.0775	0.0855
Time * School Size	6	0.002	0.0003323	1.653	0.1661	0.2365
Time * Quality Rating	3	0.001	0.0003299	1.641	0.2001	0.1883
Time * Reader	6	0.004	0.0010000	3.503	0.0092	0.0258
Error (Within factors)	31	0.006	0.0002011			

[1]P-Value corresponding to degrees of freedom adjustment using the Greenhouse-Geisser (1959) epsilon factor (= 0.692) to correct for correlation of repeated measures.

Figure Captions

Figure 1. Illustration (not to scale) of a sequence of frames in a photo-pass over a dolphin school showing image overlap between frames. Image overlap made it possible to photograph schools that occupied an area too large to capture in one photographic frame. To avoid erroneously recounting individual dolphins, readers counted, plotted and then tracked individual dolphins on adjacent frames.

Figure 2. School size frequency distribution for schools photographed during 1987 and 1988 ETP dolphin population surveys. Schools were categorized into small, medium, and large sizes. Two schools within each size category were selected for the "standard sample" used in the repeated measures experiment.

Figure 3. Comparison of mean school size and coefficient of variation for schools (n = 48) photographed during the 1989 ETP dolphin population survey.

Figure 4. Relationship between average photo quality ratings and coefficient of variation for schools (n = 48) photographed during the 1989 ETP dolphin population survey.

Figure 5. Frequency distribution of percent deviation (PD) values for each reader.

Figure 6. Percent deviation (PD) values regressed on "time." Time represents the chronological sequence, unique for each reader, in which photographed schools were counted during the counting period.

Figure 7. Temporal profiles of the mean for counts (log transformed) made by three independent readers during the repeated measures experiment.

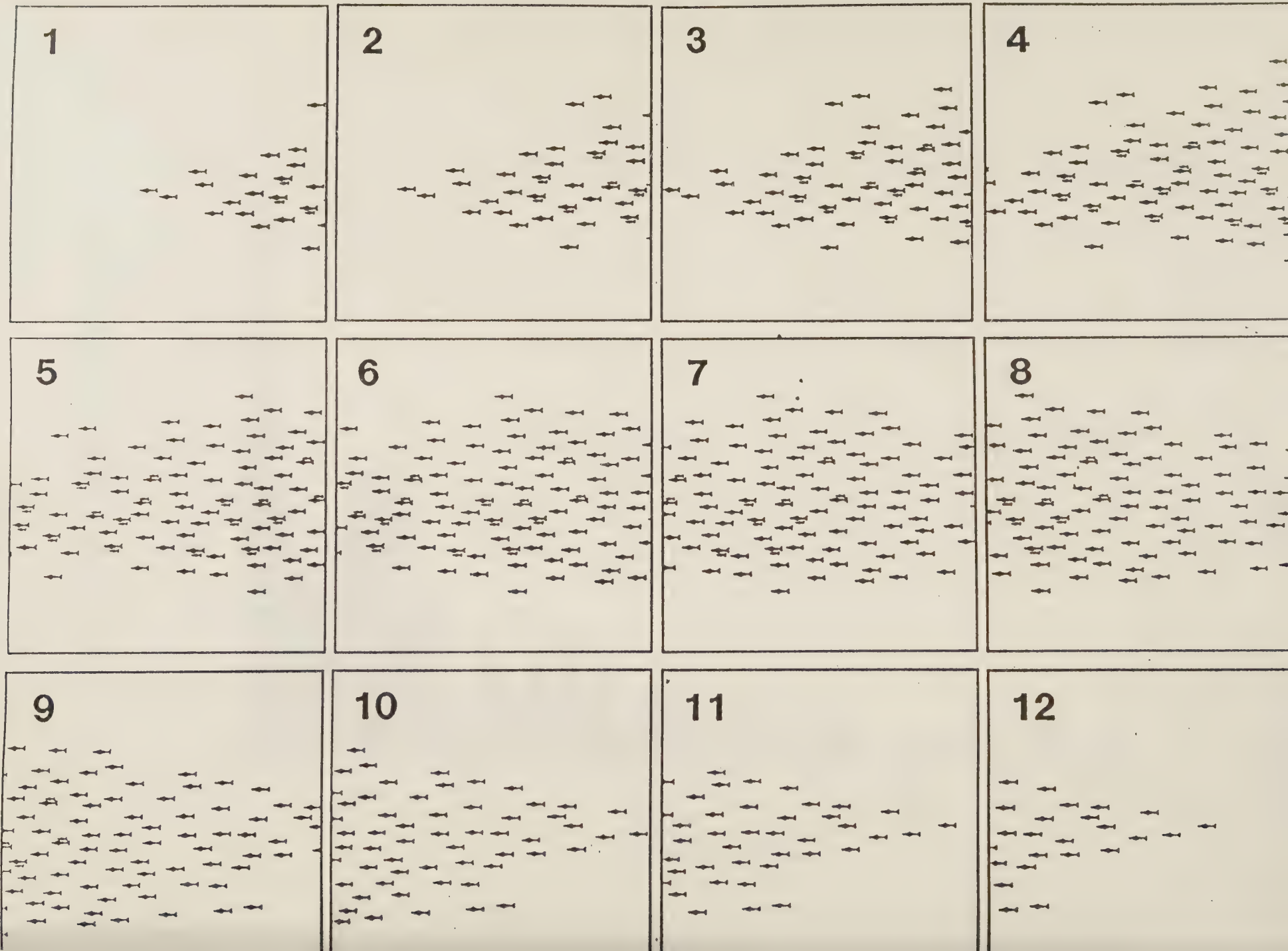


Figure 1.

Figure 2.

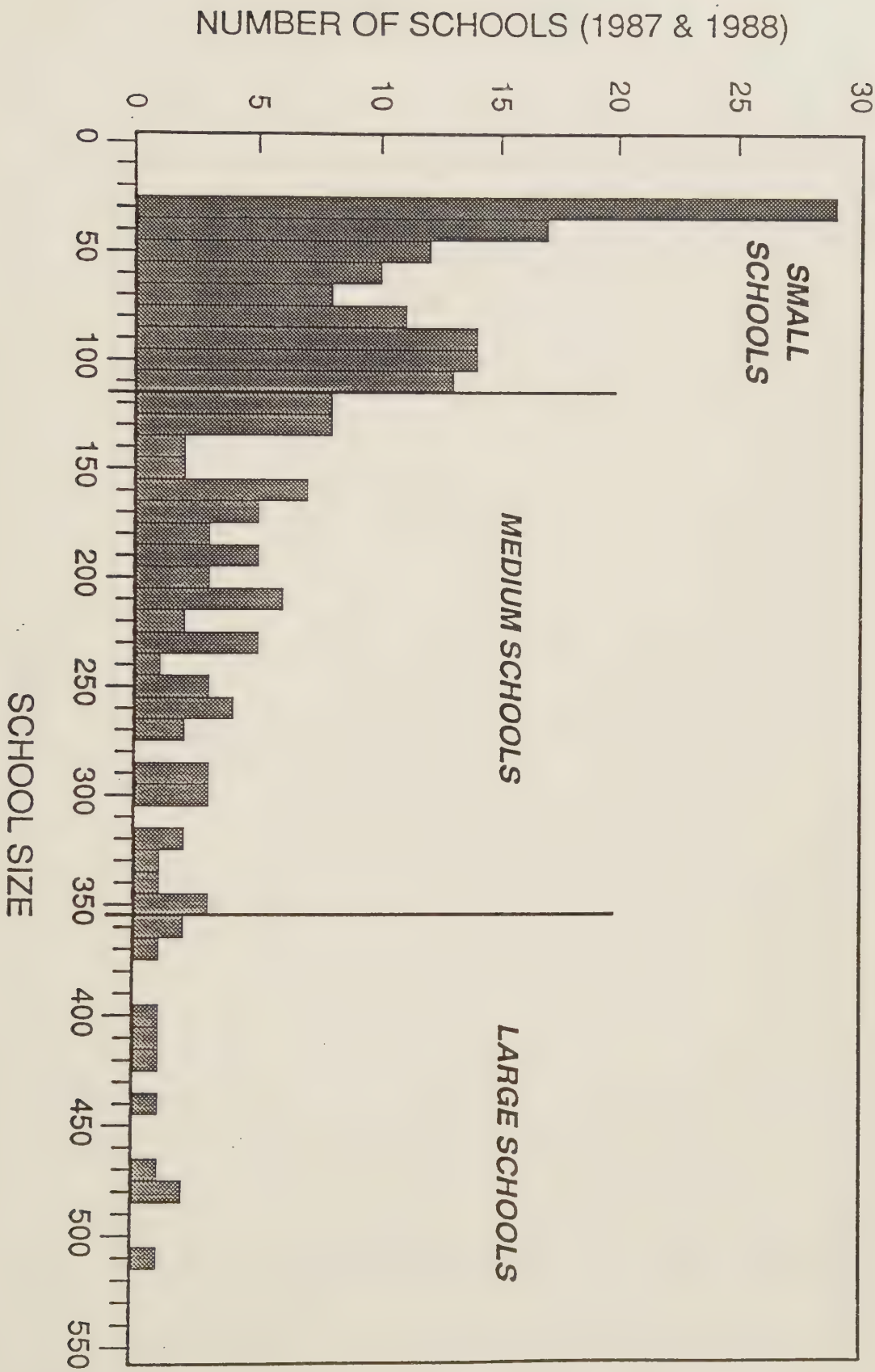

NUMBER OF SCHOOLS (1987 & 1988)
0
5
10
15
20
25
30
SMALL SCHOOLS
MEDIUM SCHOOLS
LARGE SCHOOLS
0
50
100
150
200
250
300
350
400
450
500
550
SCHOOL SIZE

Figure 3.

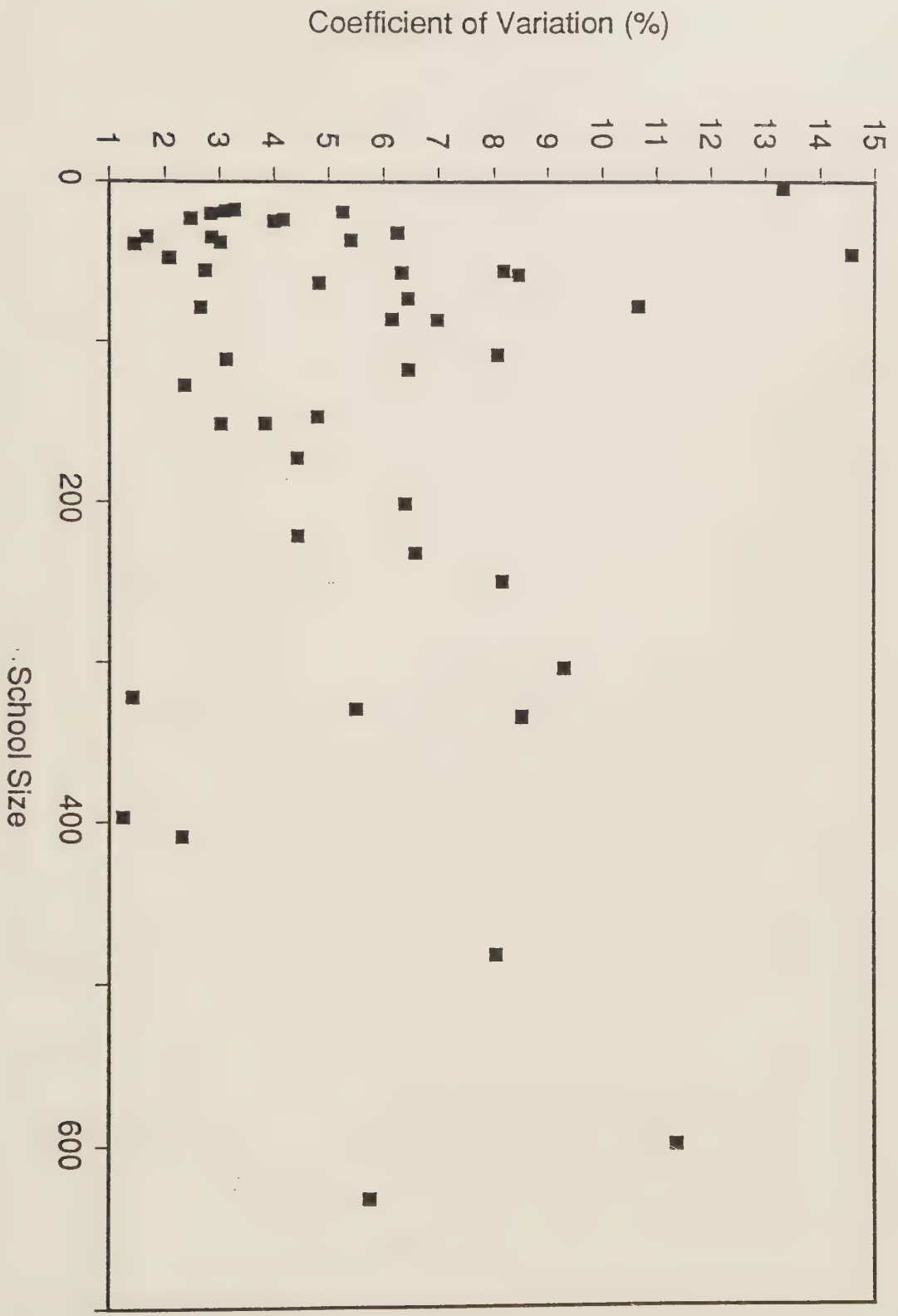

Coefficient of Variation (%)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
School Size
0
200
400
600

Figure 4.

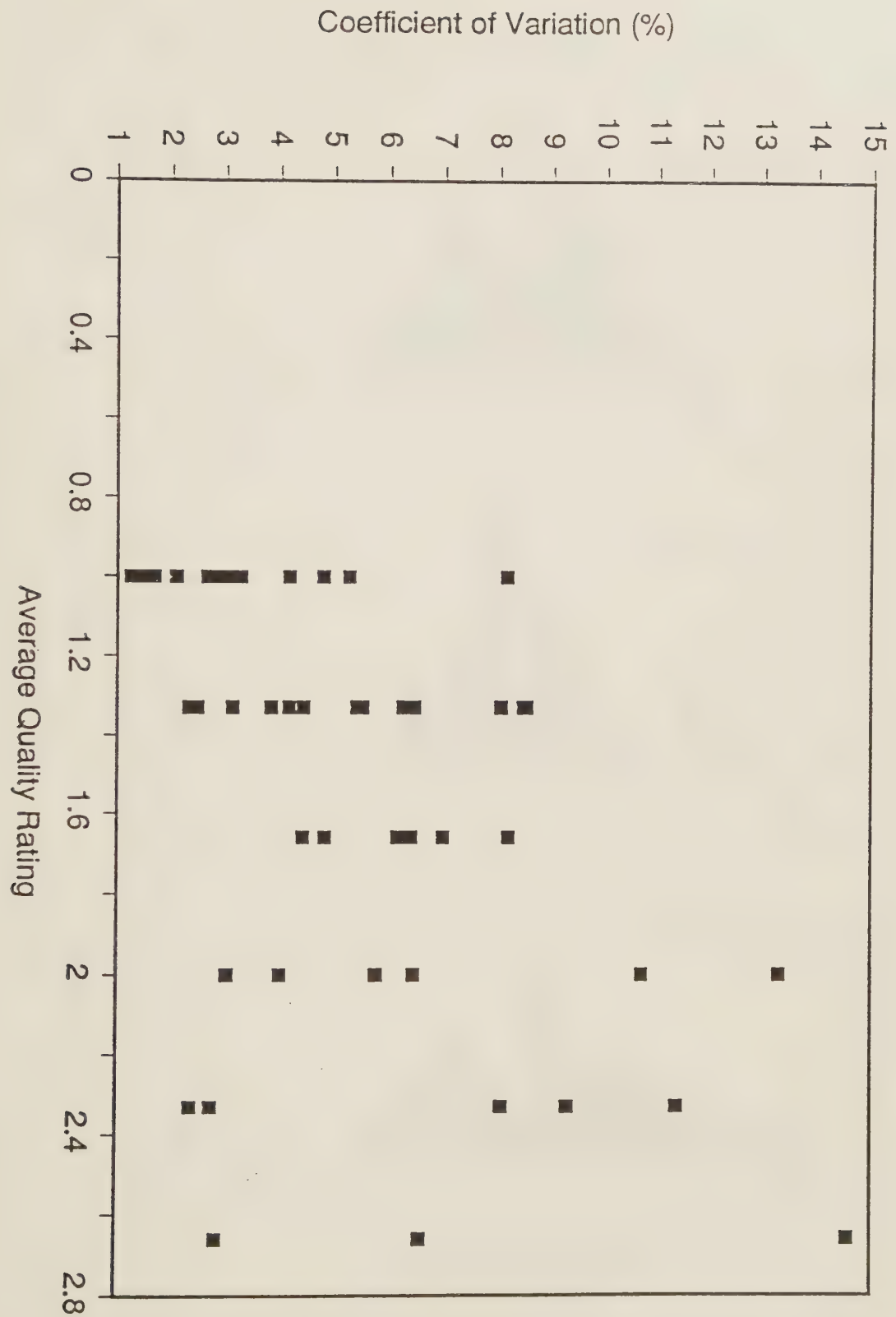

Coefficient of Variation (%)
1
2
3
4
5
6
7
8
9
10
11
12
13
14
15
Average Quality Rating
0
0.4
0.8
1.2
1.6
2
2.4
2.8

Figure 5.

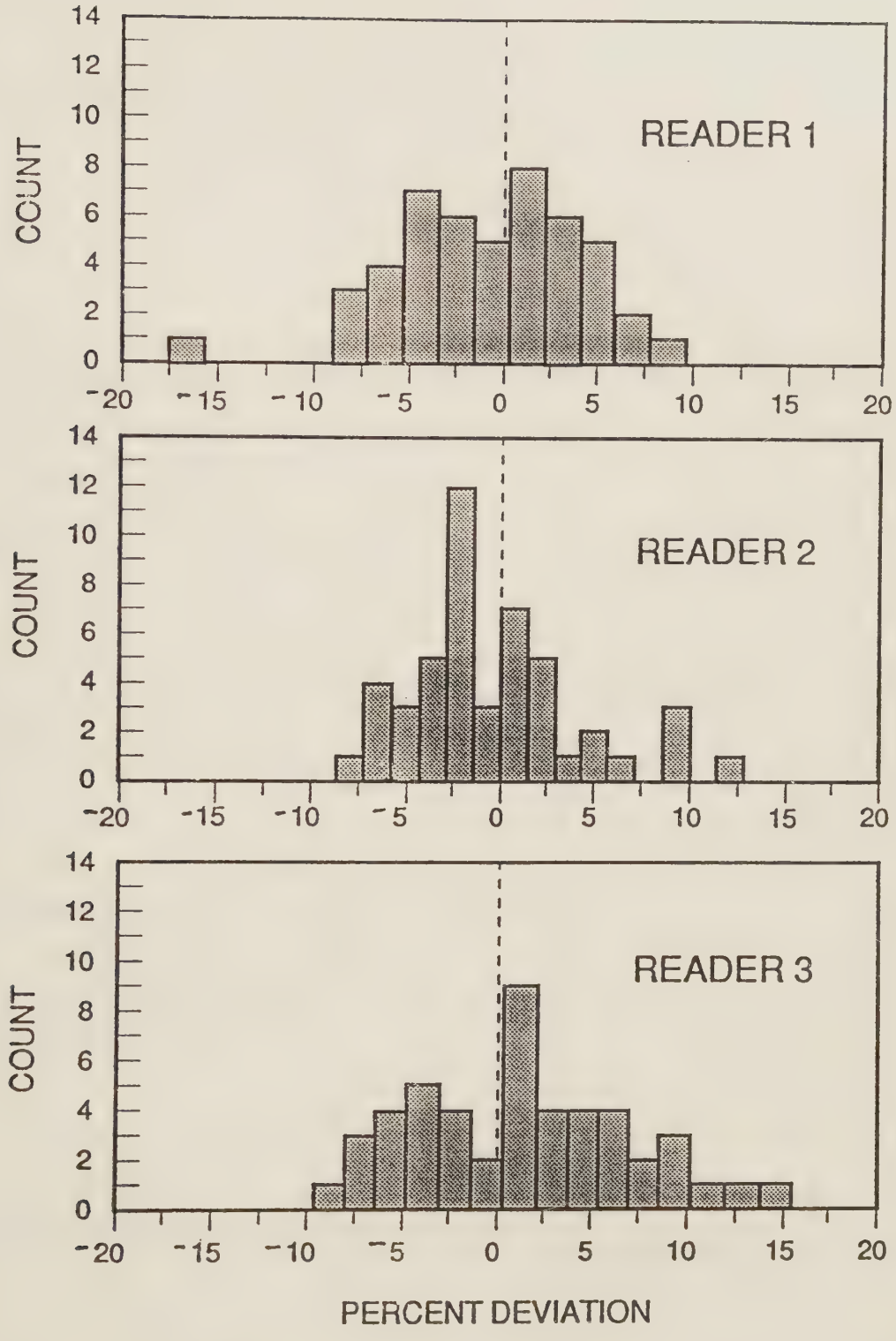

READER 1
READER 2
READER 3
COUNT
COUNT
COUNT
0
2
4
6
8
10
12
14
-20
-15
-10
-5
0
5
10
15
20
PERCENT DEVIATION

Figure 6

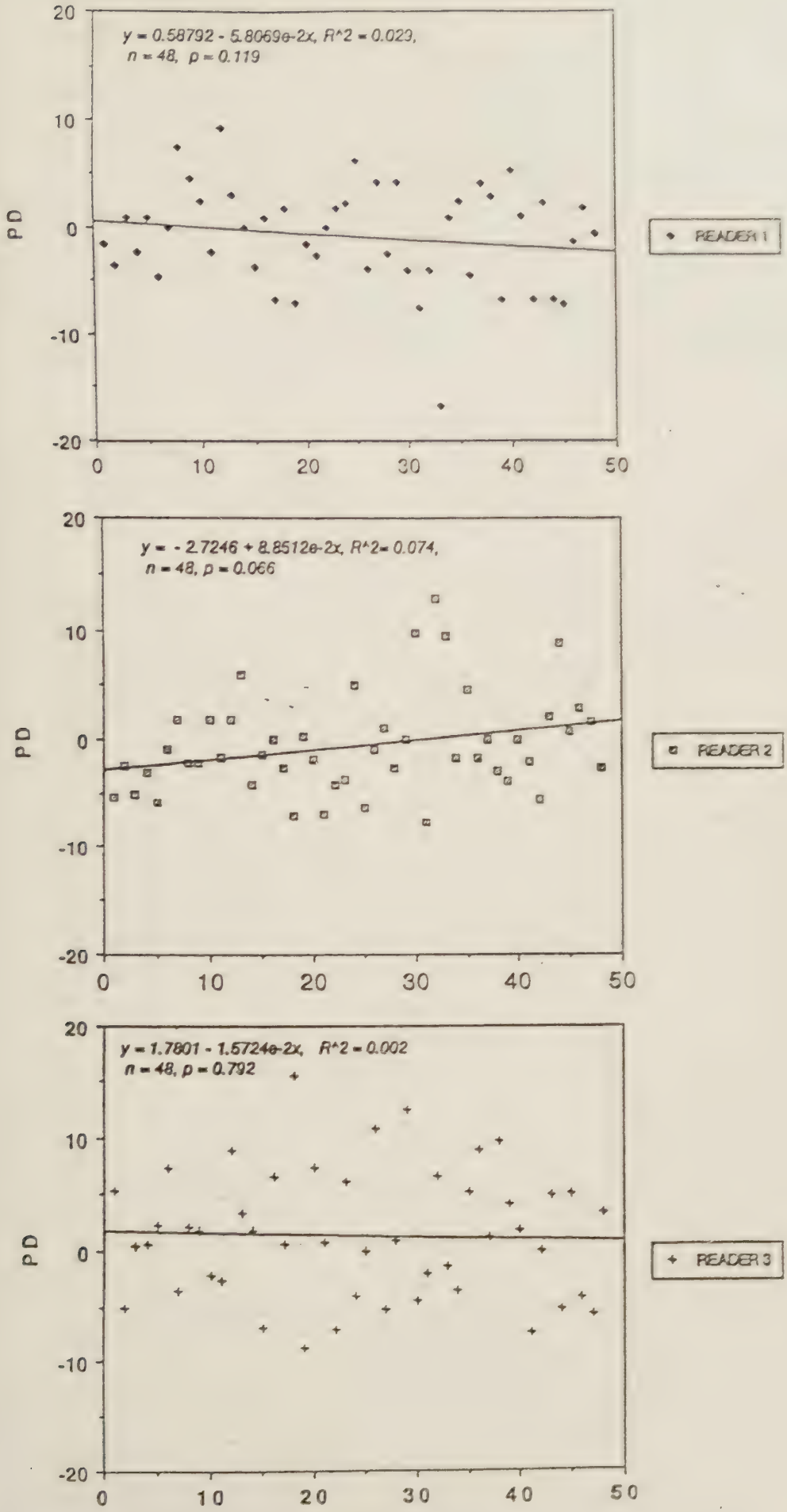
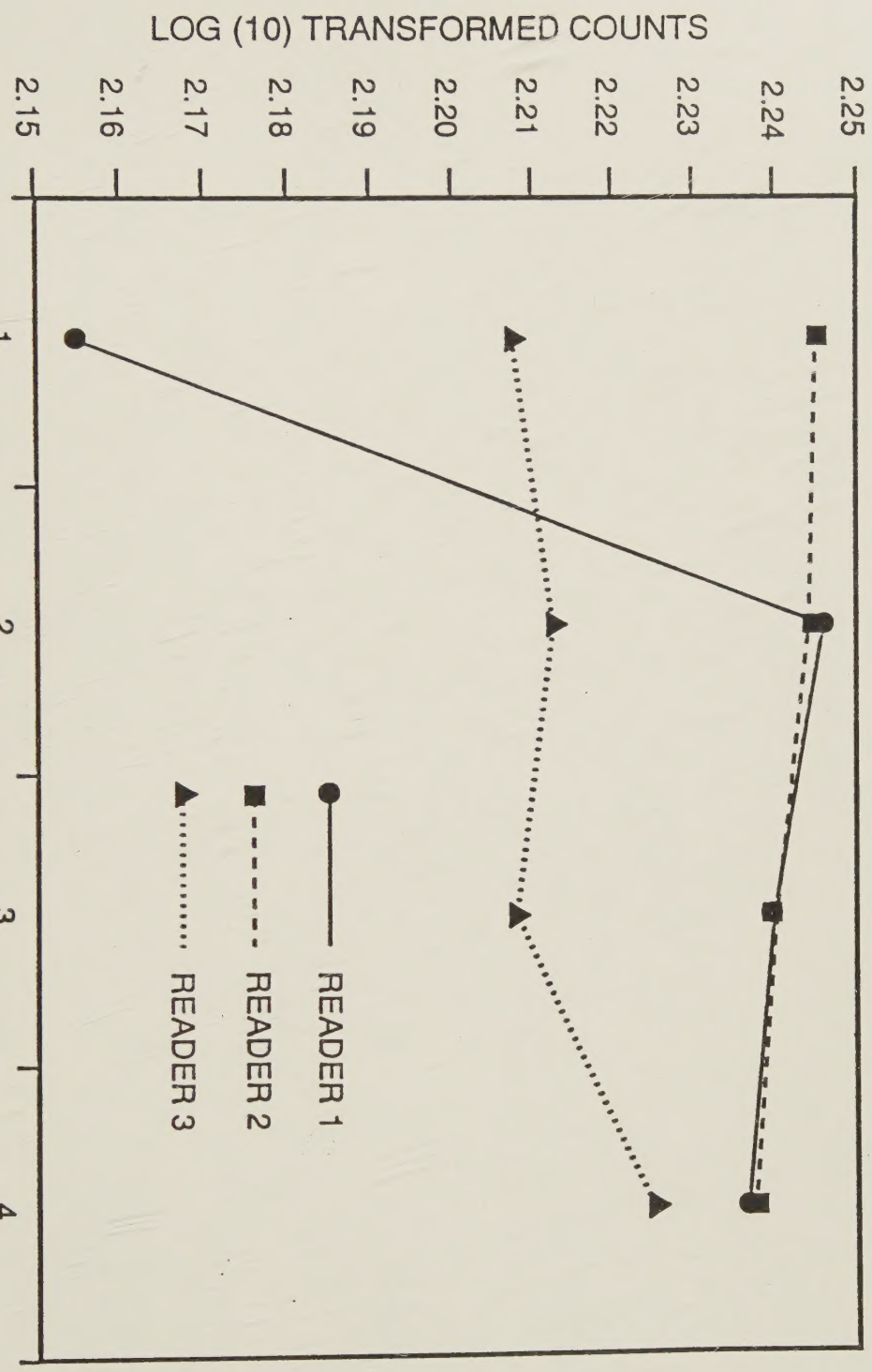

y = 0.58792 - 5.8069e-2x, R^2 = 0.029,
n = 48, p = 0.119
PD
READER 1
y = - 2.7246 + 8.8512e-2x, R^2= 0.074,
n = 48, p = 0.066
PD
READER 2
y = 1.7801 - 1.5724e-2x, R^2 = 0.002
n = 48, p = 0.792
PD
READER 3
TIME (Chronological sequence of photo counts)

Figure 7.


LOG (10) TRANSFORMED COUNTS
2.25
2.24
2.23
2.22
2.21
2.20
2.19
2.18
2.17
2.16
2.15
1
2
3
4
EXPERIMENT TIME POINTS
READER 1
READER 2
READER 3